# Memorandum

**To:** FAA ACB-330 Internal Archive

**From:** Mike Paglione, *ACB-330*

**Date:** 9/8/2002

**Re:** **Accuracy Metrics Paper for Conflict Probe Testing**

---

Attached is a draft paper in a two-column popular conference proceeding format (DASC 2002). The paper describes the accuracy metrics developed for testing the conflict predictions of an air traffic conflict probe. The paper focuses on the metrics used in the Formal Accuracy Test of URET CCLD in September 2001 for single site mode and later March 2002 for inter-facility mode. The Free Flight Office (AOZ-200) has not cleared the attached paper for publication. ***Therefore, the draft paper is strictly available for internal FAA use only***. Thank you.

# ACCURACY METRICS FOR CONFLICT PROBE TESTING

*Mike M. Paglione, Federal Aviation Administration*

*Hollis F. Ryan, and Shurong Liu, Signal Corporation*

*William J. Hughes Technical Center, Atlantic City International Airport, NJ 08405*

## Abstract

The Federal Aviation Administration's (FAA) ground based strategic conflict probe is the User Request Evaluation Tool / Core Capability Limited Deployment (URET CCLD). The prototype was developed by MITRE Corporation's Center for Advanced Aviation System Development, and the production version has been built and deployed by Lockheed Martin Air Traffic Management Division (LMATM). The FAA Formal Accuracy Testing Program was a collaborative effort between the FAA's Simulation and Analysis Group (ACB-330)[1], LMATM, MITRE, and AST Engineering Services. The formal accuracy testing measured the trajectory accuracy, conflict (i.e. loss of separation) prediction accuracy, and conflict notification timeliness of the production system and ensured these measurements met the system accuracy specifications. These requirements were distilled from the controller accepted prototype version.

This paper presents the conflict prediction accuracy metrics as applied in the Formal Accuracy Testing of URET CCLD. It describes the methodology of generating a time shifted air traffic scenario with induced aircraft-to-aircraft conflicts and encounters, the detailed process of evaluating Missed and False Alerts, and the calculation of the corresponding error probabilities. A detailed flight example is presented which illustrates the processing involved in conflict accuracy analysis. Finally, the approach is demonstrated on an entire sample scenario of many flights, similar to the Formal Accuracy Test of URET CCLD. The sample scenario results illustrate the influence that flight intent input and the design of URET alerts has on conflict prediction accuracy.

[1] The team responsible for this work, the Conflict Probe Assessment Team (CPAT), formerly was part of ACT-250.

## Introduction

In the United States, the overall system of managing and controlling air traffic is known as the National Airspace System (NAS), which is administered by the Federal Aviation Administration (FAA). Detailed procedures involving restrictions on routing, speeds, and altitudes are an integral part of the NAS. These restrictions severely reduce the amount of aircraft traffic that NAS can accommodate. A major FAA goal for improving the NAS is Free Flight. Free Flight is an air traffic control concept that increases the efficiency of aircraft operations while maintaining safety. This is achieved by introducing technology that both improves safety and allows for reductions in the restrictions imposed by the current NAS. Thus, broad categories of advances in ground and airborne automation are required. One of the most important ground based tools is a strategic conflict detection tool or conflict probe (CP). A conflict probe is a decision support tool that provides the air traffic controller with predictions of conflicts (i.e., loss of minimum separation between aircraft) for a parameter time (e.g. 20 minutes) into the future. In contrast to the current, more tactical methods of air traffic control, a conflict probe aids the controller in the strategic planning of aircraft separation management.

The User Request Evaluation Tool (URET), developed by MITRE Corporation's Center for Advanced Aviation System Development (CAASD), is a strategic conflict probe. As a result of its success and controller acceptance in Indianapolis and Memphis Air Route Traffic Control Centers (ARTCCs) as an operational prototype, the FAA contracted Lockheed Martin to build and deploy a production version of URET (known as URET Core Capabilities Limited

Deployment, CCLD) to seven ARTCCs and now has plans to implement it to the remaining thirteen.

As part of the URET CCLD deployment, the FAA Free Flight Office tasked the Simulation and Analysis Group (ACB-330) at the FAA William J. Hughes Technical Center (WJHTC) to provide input air traffic scenarios, help develop the accuracy metrics, and monitor the formal accuracy testing of the system. The testing involved measuring the URET CCLD trajectory and conflict prediction accuracy. The actual design of these metrics was a collaborative effort between ACB-330, Lockheed Martin, MITRE CAASD, and AST Engineering Services. The FAA philosophy for the Formal Accuracy Test was that URET CCLD must perform as well as, or better than, the MITRE prototype version. In September 2001, URET CCLD passed the Formal Accuracy Test in single site mode and later in March 2002 passed the test under inter-facility operation [1].

## Description of a Conflict Probe

A conflict probe is responsible for predicting into the future both the path an aircraft will fly and potential conflicts the aircraft will have with other aircraft or with restricted airspace. As implemented in URET CCLD and illustrated in Figure 1, the aircraft's trajectory and any conflict predictions are based on the flight information and track data (i.e. smoothed radar surveillance reports) from the Air Route Traffic Control Center's (ARTCC) Host Computer System (HCS), weather forecasts from the National Weather Service, and detailed adaptation databases. The databases include aircraft modeling information and system information relating to the airspace and procedures. The conflict probe uses the flight intent and tracked position information received from the HCS to build and maintain an aircraft trajectory that predicts the flight path of the aircraft. This process includes monitoring the tracked position compared to the trajectory and rebuilding it when necessary. The key element in maintaining a trajectory is that the original predicted path or trajectory is changed as more information becomes available. For example, a controller amended flight plan is received and the aircraft's trajectory is updated to match the current expanded route. By using these trajectories for all the active aircraft, the conflict probe predicts future conflicts with other aircraft and restricted airspace [2].

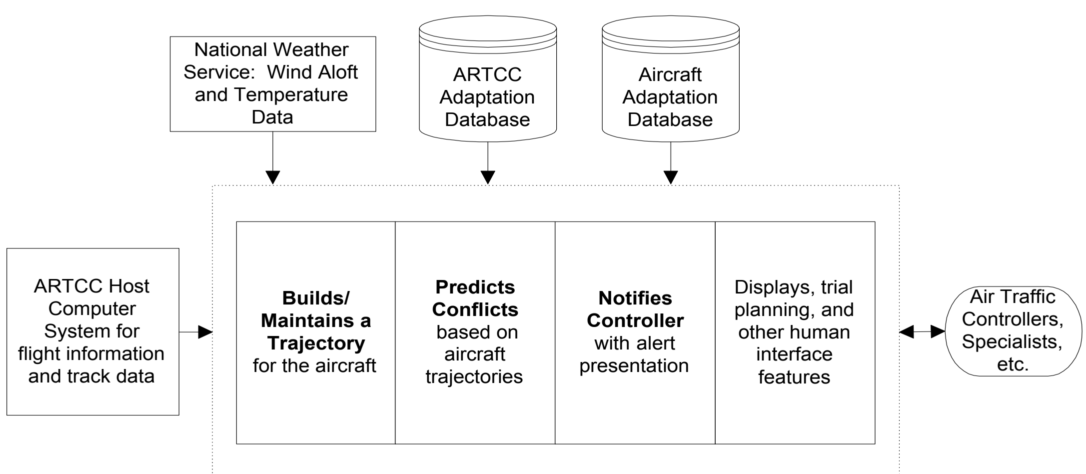


**Figure 1: Components of a Conflict Probe's Processing [2]**

## Accuracy Testing Methodology

The accuracy testing focused on three main areas of measurement:

- trajectory accuracy,
- conflict prediction accuracy,
- and conflict notification timeliness.

A conflict probe uses its predicted trajectories to determine future separation violations, i.e., to predict conflicts. Thus, the trajectory accuracy, or the deviation between the predicted trajectory and the actual path of the aircraft, has a direct effect on the accuracy of the conflict prediction. Conflict prediction accuracy is measured by several error probabilities that are used to quantify whether a predicted conflict actually occurred, and whether an actual conflict was predicted. The conflict predictions must not only be accurate in terms of the existence of a separation violation, but the conflict needs to be predicted in a timely manner. Conflict notification timeliness quantifies the amount of lead-time the probe provides in the conflict predictions.

To apply these accuracy metrics, a set of input test scenarios was generated. The test scenarios were assembled to be representative of the air traffic that URET would confront in the field. The URET systems were run without operators in real time and the output data was recorded for analysis. The CP alerts provided by the conflict probes were matched with the actual conflicts in the scenarios. Statistical tests determined whether or not, within a certain confidence, the production version performed as well as or better than the prototype [3].

This paper focuses only on conflict prediction accuracy metrics. It will describe the metrics, their rationale and method of determination, and their application on an individual flight and with a test scenario of many flights. A detailed description of trajectory accuracy is presented in References [4] and [5]. Conflict notification timeliness will be left for description in a future paper.

## Test Air Traffic Scenarios

As described earlier, test air traffic scenarios were generated as input into the production and prototype conflict probes. For acceptance testing it is important to cover all of the likely types of conflicts, while still providing realistic aircraft flight profiles.

Weather data and a recording of actual messages sent from the Host Computer System (HCS) to the CP was made at the Memphis Air Route Traffic Control Center (ZME). The HCS messages include (1) the flight plans and their amendments of all the IFR (Instrument Flight Rule) aircraft, (2) any interim altitude clearances, and (3) the radar position and velocity reports for every aircraft. Since the air traffic controllers ensure the aircraft are separated, there are no aircraft-to-aircraft conflicts in the recorded scenario. Therefore, conflicts are induced by time shifting the individual flights in the recording [6].

The amount of traffic data used depends on the goals of the particular accuracy analysis. For the URET CCLD Formal Accuracy Test, a total of six scenarios of approximately five hours in duration were generated. Each of these scenarios typically contained about 1500 flights and over 100 aircraft pair conflicts.

## Conflict Prediction Accuracy

The measurement of the accuracy of a conflict probe's predictions of aircraft-to-aircraft and aircraft-to-airspace conflicts is referred to as conflict prediction accuracy.[2] This is probably the most operationally significant metric category, since the major purpose of a conflict probe is to support strategic separation management of aircraft. Conflict prediction accuracy quantifies the fundamental error probabilities that are directly related to the probe's central goal: detecting conflicts.

[2] The focus of this paper is only on the aircraft-to-aircraft conflict predictions.

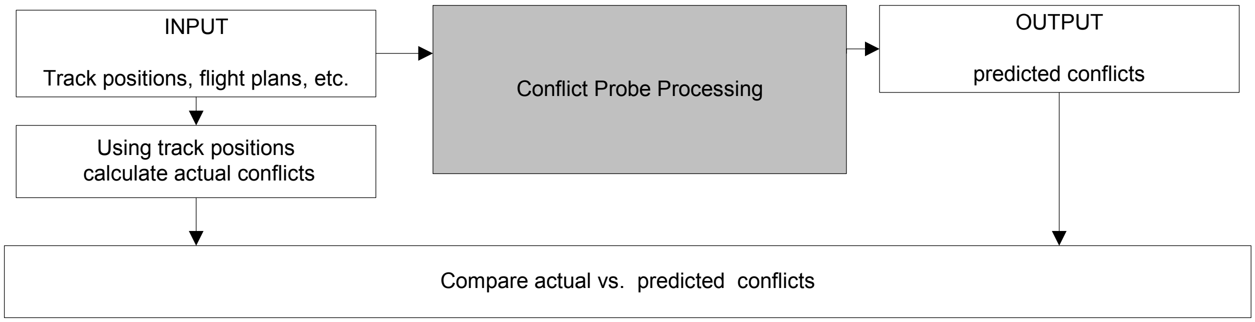


**Figure 2: General Conflict Prediction Accuracy Processing [1,7]**

In Figure 2, the conflict prediction accuracy metric isolates the conflict probe processing as a black box. Such an approach is only concerned with the input (i.e. the positions of the aircraft) and the output (i.e. predicted conflicts). A post-processing tool must first determine the actual conflicts using the aircraft position data, and then these conflicts are compared to the predicted conflicts.

### *Aircraft-to-Aircraft Conflicts and Encounters*

Once the traffic scenarios are generated the HCS track positions are checked for reasonableness and processed for determination of aircraft pair conflicts and encounters. A conflict or encounter between two aircraft occurs when their separation drops below established minima [8]. In en route airspace, while operating under IFR, aircraft are required to be at least five nautical miles horizontally separated or vertically separated by at least 1000 feet up to and including Flight Level (FL) 290, and by 2000 feet above [9]. In this paper, aircraft that violate these standard separations are considered in conflict.

It is also necessary to consider aircraft that approach each other but do not violate separation standards. A close approach not close enough to be a conflict is called an encounter. An aircraft pair, which is less than 30 nautical miles horizontally and vertically less than 4000 feet up to and including FL 290 and 5000 feet above, are considered encounters in this analysis.

ACB-330 software tools process the scenario and generate a relational database of the aircraft-to-aircraft conflicts and encounters. The fields consist of the aircraft pair's identification codes, start and end times, and other attributes of the conflict. Conflict attributes include horizontal and vertical minimum separations, vertical phase of flight, adherence age (i.e. how long before the conflict started both aircraft adhered to the HCS clearances), and pop-up category. These pop-up categories are used to excuse conflict predictions that are notified late.

The resulting conflicts and encounters generated from the scenario are used to test the conflict probe. For the accuracy testing of URET CCLD, a minimum sample size of conflicts and encounters were needed to perform the various statistical tests [1,3,7].

### *Fundamentals in Evaluating Alerts*

When the CP predicts that a future conflict will occur between two aircraft, it posts an alert to the Air Traffic Controller's display. The alert remains posted until the conflict is past or is no longer predicted. Usually the controller will redirect one of the aircraft so that the conflict will not occur. The CP automatically reads this change in flight path and deletes the alert.

The alert may be updated (in time and/or space), while it is posted to the controller's display. The initial posting of the alert and its final deletion form a notification set which can be matched to an actual conflict.

As documented in References [2,10,11,12], the CP is not perfect – it does make mistakes. For example, it can miss a conflict (Missed Alert) or it can predict a conflict that never occurs (False Alert). The four possible situations are shown in Table 1.

**Table 1: CP Alert and Conflict Event Combinations [2,12]**

| | CONFLICT OCCURS | CONFLICT DOES NOT OCCUR |
|---|---|---|
| ALERT | CP predicts conflict and it occurs (VA -- valid alerts) | CP predicts conflict and it does not occur (FA -- false alert) |
| NO ALERT | CP does not predict conflict and it occurs (MA -- missed alert) | CP does not predict conflict and it does not occur (NC -- correct no-calls) |
| Total Number of Alerts | Total Number of Conflicts | Total Number of Non-Conflicts (Encounters that did not have conflicts) |

For a real time system, it is important that an alert be given sufficiently earlier in time of the actual conflict so corrective action can be taken. In other words, an alert must be timely as well as accurate. Under normal conditions in the Formal Accuracy Test, the FAA's strategic conflict probe, URET CCLD, was required to have a five-minute lead-time or actual warning time.

As summarized in Table 1, a notification set is evaluated as a Valid Alert when CP correctly predicts the conflict and when it is posted in a timely manner. If the notification set is not presented at all or correctly predicts the conflict but is not posted soon enough, it is called a Missed Alert. The lateness of the alert may be excused only if the conflict is considered a pop-up, which is defined in detail in the later Section *Definition of Pop-Up Conflicts*. A notification set determined to be a Missed Alert due to lateness is also referred to as a Late Missed Alert or Strategic Missed Alert. A notification set presented late but excused is referred to as a Late Valid Alert.

A notification set that predicts a conflict when no conflict occurs is a False Alert. However, a False Alert withdrawn before the predicted conflict start time is also called a Retracted False Alert. A False Alert is not matched to a conflict but an encounter and may be excused as well.

Simply counting the number of times each of the events occur for a suitable mix of aircraft conflicts is not possible. It is necessary to match the alerts to the actual conflicts. There may be multiple conflicts between two aircraft. This occurs when the two aircraft are flying on close, nearly parallel paths and move in and out of conflict. Similarly there may be multiple alerts generated by the conflict probe for the same aircraft pair.

The test scenario and limitations of the conflict probe introduce additional complications. The scenario recording has a specific start time and end time. Alerts that span the start time or end time have to be treated as special cases. Also, the radar track data may be missing at the predicted conflict location. There are other considerations when input flight intent is in error. Adjustments are made for the inability of any conflict probe to predict future actions of controllers. Therefore, what appears initially to be a simple and straightforward analysis, due to the many special cases and limitations of the test scenarios and the conflict probe, ends up being quite complicated.

Taking all these factors into account, the best way to present the methodology of measuring the conflict prediction accuracy is to describe the specific process used to quantify these error events. First it is necessary to provide some definitions of key concepts. In the next two sections, adherence age and pop-up conflicts will be defined. Finally, in the last two sections, the conflict prediction processing and the error probabilities will be described.

### *Definition of Adherence Age*

Adherence is a technique to filter out conflict probe accuracy data with erroneous flight intent. Using a concept called adherence age, both Missed and False Alerts may be discarded. This can occur if the associated flights are lacking flight intent data at either the Missed Alert's conflict start time or False Alert's predicted conflict start time.

Normally the pilot of an aircraft flies along an approved route (i.e. entered into the HCS and thus known to the CP), which is usually from navigation aid to navigation aid. However, sometimes the pilot either strays from the approved route or is cleared only verbally by the controller. In this analysis, the aircraft is considered to be out of adherence if it strays beyond a set of thresholds either laterally or vertically from its HCS entered clearances.

As defined in detail in Reference [13] and established in Reference [7], Table 2 lists the lateral thresholds, partitioned on whether the track is associated with a turn or straight portion of the route and the reported altitude. For level flight, the vertical threshold is 300 feet below FL 290 and 500 at and above FL 290. When the flight is in vertical transition (i.e. climbing or descending), the track is assumed to be in vertical adherence.

**Table 2: Lateral Adherence Thresholds [13]**

| Associated Altitude (A in 100's of feet) | En route Threshold (nautical miles) | Turn Threshold (nautical miles) |
|---|---|---|
| A ≤ 100 | 13 | 11 |
| 100 < A ≤ 180 | 16 | 13 |
| 180 < A ≤ 330 | 19 | 13 |
| 330 < A | 19 | 14 |

The adherence age is the length of time that the aircraft has been in continuous adherence[3]. It is associated with each radar track report. The adherence age of a conflict is the lesser of each aircraft's individual adherence age at the conflict start time. This is an attribute of the conflict and is only used to evaluate Missed Alerts. For Missed Alerts, the conflict's adherence age must be above a user-defined threshold (e.g. 13 and 20 minutes were used in URET CCLD Formal Accuracy Test).

The adherence age for a False Alert, when no conflict is available, is the lesser of each aircraft's individual adherence age taken at the predicted conflict start time of the corresponding notification set. A False Alert is discarded if this adherence age is less than the duration between predicted conflict start time and the notification set start time. Therefore, when adherence is applied, the potentially false conflict prediction is filtered out if during the CP's look ahead into the future one of the aircraft displayed a lack of flight intent.

In summary, the Missed and False Alerts may be excused when adherence is applied and the CP is not provided with the adequate aircraft intent information. This is achieved by checking HCS track reports laterally and vertically against the current HCS clearances and measured in time using the attribute called adherence age.

### *Definition of Pop-Up Conflicts*

As discussed earlier, for a CP to have strategically predicted a conflict, it must be notified at least five minutes prior to the actual conflict start time. This conflict timeliness requirement for a Valid Alert is relaxed if the conflict is considered a pop-up. A pop-up conflict occurs if the CP is not provided with same five-minute time threshold of continuous HCS data or prediction for either of the associated flights. As documented in Reference [14], there are six different reasons for a conflict being labeled a pop-up, and they include:

1. The conflict starts within five minutes of the start of either aircraft's HCS track. For example, this occurs when the conflict starts as one of the associated aircraft enters the scenario.
2. The conflict starts within five minutes of a HCS clearance.
3. The conflict starts within five minutes from the time either aircraft exit an inhibited airspace not modeled for aircraft-to-aircraft conflicts. For URET, these airspace boundaries usually include terminal areas where separation rules differ from en route airspace.
4. The conflict starts within five minutes of either aircraft having a gap in track data, which is greater than 2 minutes.
5. The conflict occurs when either aircraft is less than an adapted altitude (e.g. 300 feet) from a cleared interim or hold altitude at the conflict start.
6. The conflict starts within five minutes from when the associated alert is updated from being muted in color. URET presents some alerts as muted (i.e. level 2) during a portion of flight not

[3] If in adherence, as defined above, a track report's adherence age defaults to infinity at the start of HCS track or following a flight plan amendment. It also defaults to zero following a gap in HCS track reports of 2 minutes or more.

yet vertically cleared by air traffic control. This will be explained in more detail in the later Section *Sample Scenario Analysis with URET*. Unlike the other five cases, this particular pop-up is URET specific, since it involves URET's unique method of muting the color designations of its alerts.

These situations allow relaxation of the Valid Alert conflict timeliness requirement, since under these conditions a strategic conflict probe would not be expected to predict the conflict beyond the five-minute threshold. However, regardless whether the conflict is a pop-up, a Valid Alert still needs to be posted prior to the actual conflict start.

## *Methodology of Conflict Prediction Accuracy*

The Missed, Valid, and False Alerts, as defined in Table 1, are determined in two sub-processes. In Process A (see Figure 3), conflicts are evaluated in order of actual conflict start time and matched against eligible notification sets. To even be eligible for matching to a specific conflict, a notification set must have a posting time prior to the start of the conflict and must have an end or delete time after the start of the actual conflict. Thus, the notification must precede the conflict and must be active at the start of the conflict. The result is a new listing of Valid Alerts, Missed Alerts, and discarded conflicts.

The discarded conflicts are either conflict pairs that have no eligible notification set or were notified late (i.e. a Strategic Missed Alert) but can be discarded with low adherence age. For Process A, only lack of adherence can excuse a Missed Alert, while adherence is not even checked if a Valid Alert is determined. In other words, if a CP correctly predicts a conflict, the aircraft pair's flight intent is irrelevant, allowing a CP with superior heuristics that handle out of adherence situations more effectively to achieve higher accuracy results.

The remaining notification sets not matched as either Valid Alerts or discarded conflicts are potentially False Alerts. In Process B (see Figure 4), the remaining notification sets are evaluated to determine which of them are truly False Alerts and which can be discarded. Unlike the Missed Alerts, there are several reasons for discarding False Alerts. The potential False Alert is discarded if either aircraft does not have HCS track data present at the predicted conflict start time (PCST). With a lack of HCS track data, the False Alert error is unverifiable and thus excused. In many of these cases, the discarded notification sets represent alerts predicted beyond the end of the traffic scenario.

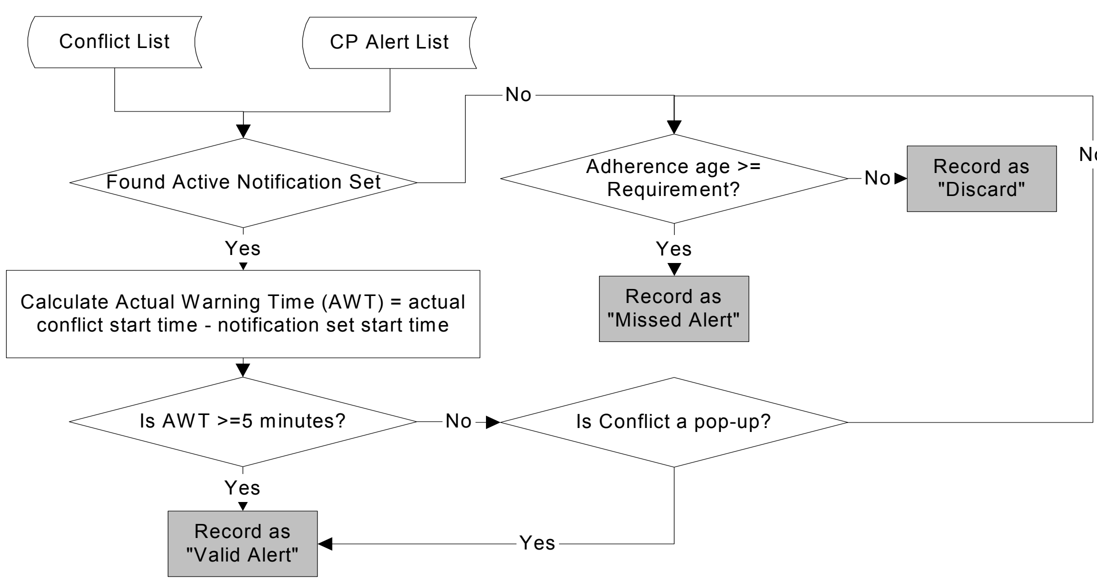


**Figure 3: Process A – Valid and Missed Alert Processing**

The potential False Alert is discarded if either aircraft has a low adherence age at the predicted conflict start time. As discussed in the previous Section *Definition of Adherence Age*, this notification set may be discarded, when the involved aircraft have inadequate flight intent information available within the planning horizon of the prediction.

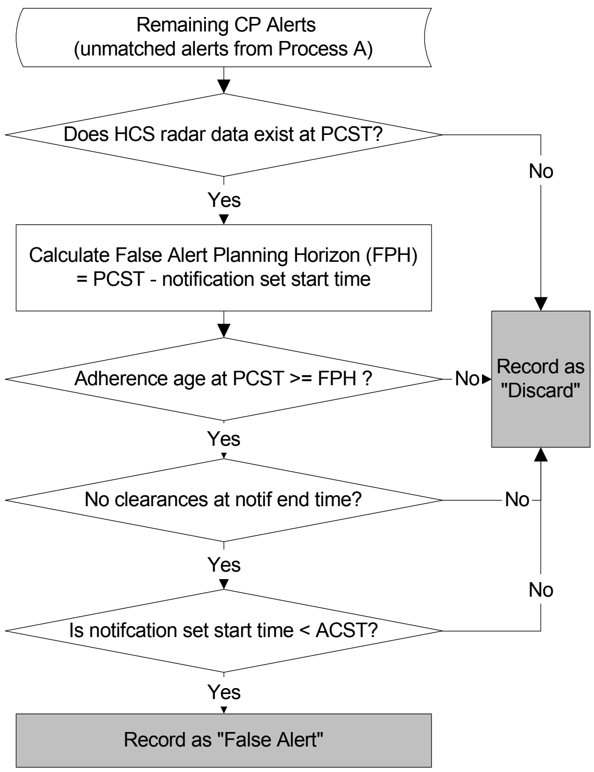


**Figure 4: Process B – False Alert Processing[4]**

If the potential False Alert is retracted due to an air traffic control clearance, the notification set is discarded. The potential False Alert can also be discarded if the notification set was posted after the last actual conflict start time (ACST) between the associated aircraft. This can only happen if a conflict actually occurs between these aircraft and another alert is presented after it starts. When the CP is operating in the NAS, once the actual conflict started, strategic alerts would have little value and other more tactical procedures would be utilized.

### *Probability Definitions*

The Missed and False Alerts counts are normalized by dividing them by the number of conflicts and encounters they are matched to. The resulting ratios are the probability of Missed and False Alerts. Equation 1 defines the probability of Missed Alert. It quantifies the conditional probability that the CP does not predict the conflict when it occurs.

$$P(MA) = \frac{MA}{(VA + MA)}, \qquad \textbf{Equation 1}$$

where $MA$ is the number of Missed Alerts and $VA$ is the number of Valid Alerts.

The False Alert probability is defined as the likelihood in predicting a conflict when it does not occur. This is defined in Equation 2. False Alert probabilities are partitioned by the horizontal separation of their corresponding encounters. Table 3 presents a sample of the bins used for the accuracy acceptance testing of URET CCLD. Therefore, the False Alert probability is further conditioned on the horizontal minimum separation. For example, for Bin 1 in Table 3, the metric is defined as the probability of a falsely predicting an encounter between aircraft with less than 10 nautical miles horizontal separation as a conflict.

$$P(FA) = \frac{FA_i}{E_i}, \qquad \textbf{Equation 2}$$

where $FA_i$ is the number of False Alerts matched to encounters in Bin i and $E_i$ is the total number of encounters in Bin i found in the given traffic scenario.

**Table 3: Example FA Bins Used for URET [3]**

| Bin Number (i) | Horizontal Separation (H in nautical miles) |
|---|---|
| 1 | $H < 10$ |
| 2 | $10 \leq H < 15$ |
| 3 | $15 \leq H < 23$ |
| 4 | $H \geq 23$ |

[4] PCST is the predicted conflict start time of the notification set and ACST is the actual conflict start time of the true conflict.

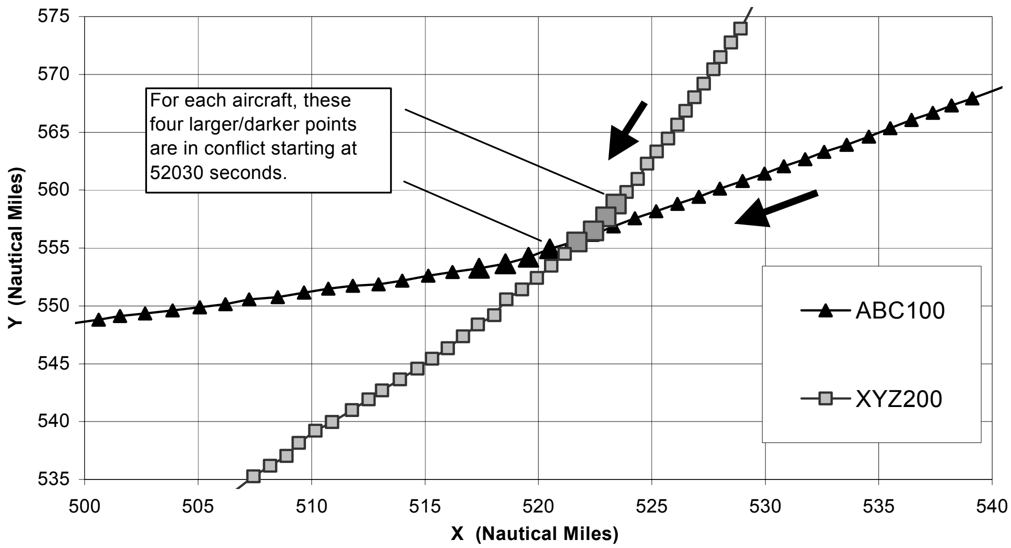


**Figure 5: Horizontal Profile of ABC100 and XYZ200 Conflict Example**

## Example Flight Analysis with URET

A flight example, referred to in this paper as ABC100, was selected from a Memphis ARTCC (ZME) test scenario. It was first presented in Reference [5] in December 2001 to illustrate how the trajectory prediction accuracy methodology is applied. The focus of this paper is on conflict prediction accuracy, so the analysis of this same flight's actual and predicted conflicts are presented. The conflict probe used for this example is the URET Prototype[5]. Flight ABC100 is an over flight, entering the ZME airspace at Flight Level 350 (FL350), descending to FL310, and then exiting the ZME airspace at this altitude. The aircraft is cleared to descend to FL310 at 14:25:05 and the resulting Top Of Descent (TOD) time is at 14:25:10 (51910 seconds). For ZME and the analysis, the flight concludes at 14:48:00 (53280 seconds), when air traffic control of ABC100 is passed to the Fort Worth ARTCC (ZFW).

In this time shifted test scenario, as shown horizontally in Figure 5, the flight XYZ200 is cruising at FL310 crosses ABC100's route at an encounter angle of 38 degrees. As shown vertically in Figure 6, this crossing encounter occurs while ABC100 is descending to FL310 causing a test conflict with a minimum horizontal and vertical separation of 4.8 nautical miles and 1050 feet, respectively. The conflict is rather short starting approximately at 14:27:10 and ending at 14:27:40. Once again, this aircraft-to-aircraft conflict is not real, but induced in the test scenario simply by time shifting the flights. However, the URET conflict probe tested with these flights is expected to predict the conflict as if it were real.

As presented in the Table 4, the URET Prototype presents four notification sets, where the first three are all retracted before the conflict started. Notification Set A was presented at 14:10:25 but was retracted only four seconds later at 14:10:29. This notification set was evaluated as a Retracted False Alert. Approximately a minute and a half later at 14:11:58 Notification Set B was presented and again retracted, producing a second Retracted False Alert. At 14:20:57 yet another notification set, Notification Set C, was presented, but it was almost immediately retracted at 14:21:00. This retraction was caused by a hold altitude clearance and consequently was discarded.

[5] MITRE developed URET Prototype system, Release URETD32R2LMP1C. It is referred to as the baseline URET prototype for URET CCLD.

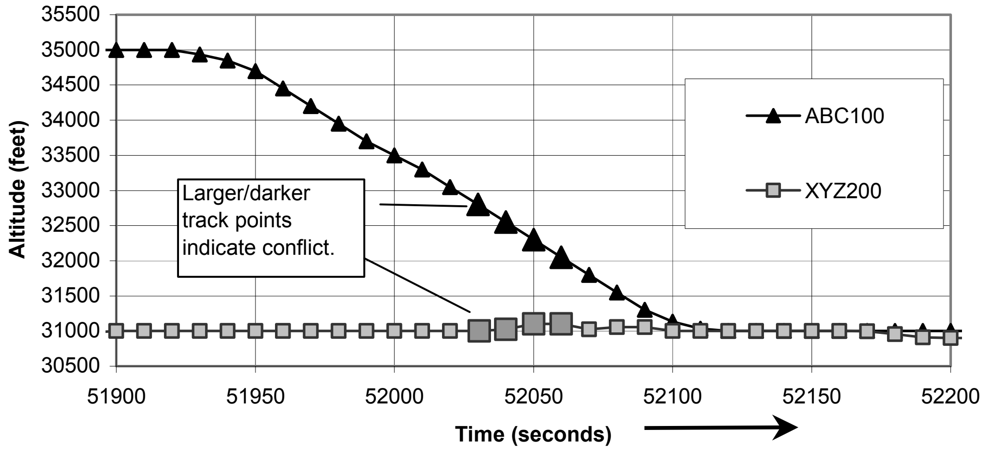


**Figure 6: Vertical Profile of ABC100 and XYZ200 Conflict Example**

**Table 4: Notification Sets for ABC100 and XYZ200 Conflict Example**

| Notification Set | Notification Start Time | Notification Set End Time | Predicted Conflict Start Time | Predicted Conflict End Time | Description |
|---|---|---|---|---|---|
| A | 14:10:25 | 14:10:29 | 14:24:35 | 14:29:47 | Retracted False Alert. |
| B | 14:11:58 | 14:14:29 | 14:26:21 | 14:30:00 | Retracted False Alert. |
| C | 14:20:57 | 14:21:00 | 14:25:24 | 14:29:50 | Retracted False Alert discarded due to clearance. |
| D | 14:25:05 | 14:29:56 | 14:26:23 | 14:29:56 | Valid Alert 2:05 before pop-up |

Finally at 14:25:05 a fourth notification, Notification Set D, was presented and remained active until the conflict started two minutes in five seconds later at 14:27:10. Thus, this last notification set is a Valid Alert matched to the ABC100 and XYZ200 conflict. As discussed in the previous *Fundamentals in Evaluating Alerts* Section, the conflict probe is normally required to present an alert five minutes before the conflict actually starts, however this is relaxed if the conflict is labeled a pop-up. The ABC100 flight was cleared to descend at 14:25:05, started its descent five seconds later, and the conflict started roughly two minutes later at 32800 feet. The conflict started within the defined five minutes of a clearance labeling it a pop-up conflict.

In summary, for this ABC100 and XYZ200 conflict the URET Prototype produced two False Alerts and one Valid Alert. It illustrates several of the conflict prediction accuracy rules. These notification sets are only a few of thousands that were evaluated in the Formal Accuracy Testing of URET CCLD. Among others, this particular aircraft pair conflict was used to validate the analysis software used for this test.

## Sample Scenario Analysis with URET

In this section, a sample scenario is reported upon that illustrates the conflict prediction accuracy methodology but is not representative of the overall performance of URET. Like the flight example, the conflict probe used for this analysis is the URET Prototype. This sample data is only about three percent the traffic quantity used in the Formal Accuracy Test. However, the sample presented will effectively demonstrate the conflict prediction methodology, which is the focus of this paper. The complete results of the URET CCLD Formal Accuracy Test are documented thoroughly in Reference [1].

The sample test scenario includes approximately 500 flights and 2 hours in duration from Indianapolis ARTCC and was time shifted to produce 211 aircraft-to-aircraft conflicts. This was accomplished using a recently developed technology by ACB-330, utilizing a random search heuristic called a genetic algorithm [15]. The resulting conflict predictions from the URET Prototype system were analyzed, using ACB-330's suite of conflict prediction accuracy tools.

URET presents its aircraft-to-aircraft conflict predictions as red, yellow, and muted red or yellow alerts. URET's red alerts are predicted to have a minimum horizontal separation of less than five nautical miles. The remaining non-muted alerts are presented yellow and are predicted to have separations less than a range of 8 to 12 nautical miles. Indicated by lighter shading of its red and yellow alerts, URET also mutes its alerts when it has predicted conflicts in a portion of flight not yet cleared vertically by air traffic control. The muting of alerts is an advanced URET feature, designed to balance the False Alerts potentially generated on an uncleared portion of a flight, yet providing the controller with as much warning of prospective conflicts as possible.

For this example analysis, both red and yellow alerts will be considered as valid conflict notifications. However, the adherence rule (i.e. lack of flight intent) and muted alerts will be contrasted on the False Alert performance. In Figure 7, the False Alert Probability is partitioned by minimum horizontal separation as described in Table 3 and Equation 2. Since red and yellow alerts are included in this example analysis and have predicted horizontal separations of up to 12 nautical miles, Figure 7 will start on Table 3's second bin (i.e. 10 to 15 nautical mile horizontal separation). This is contrasted to the Formal Test that focused only on the more operationally significant red alerts where the first bin of less than 10 nautical miles has relevance [1].

More of interest then URET's performance in a particular bin is the significant decrease in False Alert Probability as horizontal separation increases and the contrast between the effects of adherence and muting. As discussed in Reference [2], the sharpness or steepness of the alert probability curve indicates a conflict probe is performing well. In this sample scenario, URET's False Alert Probability ranges from 0.41 to 0.11 between 10 and 23 nautical miles horizontal separation. In Figure 7's second bin between 10 and 15 nautical miles, the lack of flight intent as measured by the adherence rule causes URET to approximately double the amount of probability of False Alerts. This is fairly consistent with similar conflict probes developed by NASA as presented in Reference [10]. Once again, in Figure 7's second bin between 10 and 15 nautical miles, the muting also illustrates a two-fold decrease in accuracy. Therefore, for this sample scenario the URET Prototype False Alert Probability decreases sharply, and the flight intent and muting significantly affect its accuracy measurement.

For this sample scenario, URET's Missed Alert Probability for all alerts (i.e. red, yellow, and muted) is 0.03 with the proper flight intent as measured by adherence age but 0.07 when intent is not considered. If muting was excluded as presented in Figure 7 for False Alert accuracy, the Missed Alert Probability approximately doubled. Once again, flight intent and muting logic will significantly affect conflict prediction accuracy, so need to be considered.

## Summary

This paper presented conflict prediction accuracy metrics as applied in the Formal Accuracy Testing of URET CCLD. The methodology of generating a time shifted air traffic scenario with induced aircraft-to-aircraft conflict and encounters, the two-stage process of evaluating Missed and False Alerts, and the calculation of the corresponding error probabilities are presented. A specific flight is presented which illustrates the application of the conflict accuracy measurement rules. Finally, the approach is demonstrated on an entire scenario of many flights as in the Formal Testing of URET CCLD. The sample scenario analysis further illustrates the influence that a conflict probe's flight intent input data and design features of alerts (e.g. URET's muted conflict notifications) influence its conflict prediction accuracy and need to be considered.

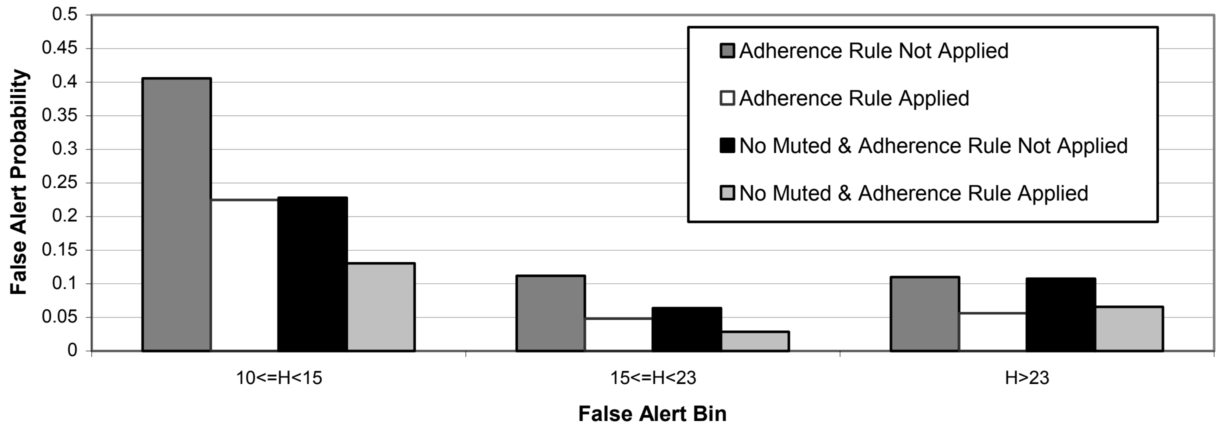


**Figure 7: Sample False Alert Probability**

## Acronyms

ACB-330 Simulation and Analysis Group at the FAA WJHTC
ACT-250 Engineering and Integration Services Branch at the FAA WJHTC
ARTCC Air Route Traffic Control Center
CAASD Center for Advanced Aviation System Development
CCLD Core Capability Limited Deployment
CP Conflict Probe
CTAS Center-TRACON Automation System
DST Decision Support Tool
FAA Federal Aviation Administration
FL Flight Level
HCS Host Computer System
IFR Instrument Flight Rules
NAS National Airspace System
NASA National Aeronautics and Space Administration
TOD Top Of Descent
URET User Request Evaluation Tool
WJHTC William J. Hughes Technical Center
ZID Indianapolis ARTCC
ZFW Fort Worth ARTCC
ZME Memphis ARTCC

## References


[1] Lockheed Martin Air Traffic Management, April 25, 2002, "User Request Evaluation Tool (URET) Core Capability Limited Deployment (CCLD) Test Report for the Formal Accuracy Test, Volume III," Lockheed Martin Corporation, Rockville, MD.

[2] Paglione, M., M. Cale, H. Ryan, Fall 1999, "Generic Metrics for the Estimation of the Conflict Prediction Accuracy of Aircraft to Aircraft Conflicts by a Strategic Conflict Probe Tool," *Air Traffic Control Quarterly*, Vol. 7 (3).

[3] Lockheed Martin Air Traffic Management, August, 1998, "User Request Evaluation Tool Core Capability Limited Deployment System Specification (SSS), Volume I, Part 2: Conflict Probe (CP)," Lockheed Martin Corporation, Rockville, MD.

[4] Paglione, M., H. F. Ryan, R. D. Oaks, J. S. Summerill, M. L. Cale, May 1999, "Trajectory Prediction Accuracy Report User Request Evaluation Tool (URET)/Center-TRACON Automation System (CTAS)," DOT/FAA/CT-TN99/10, FAA WJHTC/ACT-250.

[5] Paglione, M., R. Oaks, M. L. Cale, S. Liu, H. Ryan, J. S. Summerill, December 3, 2001, "A Generic Sampling Technique for Measuring Aircraft Trajectory Prediction Accuracy," 4th USA/EUROPE Air Traffic Management R&D Seminar.

[6] Oaks, R., M. Paglione, Fall 2001, "Generation of Realistic Air Traffic Scenarios Based on Recorded Field Data," *46th Annual Air Traffic Control Association Conference Proceedings*, Arlington, VA, pp.142-146.

[7] Paglione, M., R. D. Oaks, H. F. Ryan, J. S. Summerill, January, 27, 2000, "Description of Accuracy Scenarios for the Acceptance Testing of the User Request Evaluation Tool (URET) / Core Capability Limited Deployment (CCLD), Final," FAA WJHTC/ACT-250.

[8] Bilimoria, K. D., H. Q. Lee, August, 2001, "Properties of Air Traffic Conflicts for Free and Structured Routing," paper presented at the AIAA Guidance, Navigation, and Control Conference, Montreal, Canada.

[9] United States Department of Transportation, Federal Aviation Administration, February 24, 2000, *Air Traffic Control 7110.65M*, FAA.

[10] Bilimoria, Karl, May-June 2001, "A Methodology for the Performance Evaluation of a Conflict Probe," *Journal of Guidance, Control, and Dynamics*, Vol. 24 (3).

[11] Brudnicki, D., W. Arthur, K. Lindsay, April 1998, "URET Scenario-based Functional Performance Requirements Document," MTR98W0000044, MITRE/CAASD.

[12] Cale, M. L., M. Paglione, H. Ryan, D. Timoteo, R. Oaks, April 1998, "URET Conflict Prediction Accuracy Report," DOT/FAA/CT-TN98/8, FAA WJHTC/ACT-250.

[13] Paglione, M. M., J. S. Summerill, August 23, 2000, "Study of Lateral Adherence Thresholds for User Request Evaluation Tool Core Capability Limited Deployment Accuracy Acceptance Testing," FAA WJHTC/ACT-250.

[14] Lockheed Martin Air Traffic Management, June 1, 2001, "User Request Evaluation Tool Core Capability Limited Deployment Software Requirements Specification Conflict Probe Algorithmic Data Dictionary, Volume V," Lockheed Martin Corporation, Rockville, MD.

[15] Oaks, Robert D., August 2002, "A Study on the Feasibility of Using a Genetic Algorithm to Generate Realistic Air Traffic Scenarios Based on Recorded Field Data," paper presented at the AIAA Guidance, Navigation, and Control Conference, Monterey, CA.